

Desarrollo de un Sistema Experto para el Proyecto de Investigación “Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales”

Damiano Luis Esteban, Paz Menvielle Alejandra, Muñoz Roberto Miguel,
Romero María Soledad, Cuevas Juan Carlos, Bartó Carlos, Nicolas Fiorito,
Mauricio Spalletti

Departamento Ingeniería en Sistemas de Información / Facultad Regional
Córdoba / Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Ciudad Universitaria - Córdoba
0351 - 4686385

luis.damiano@gmail.com, pazmalejandra@gmail.com, robertmunioz@gmail.com,
romeroma.soledad@gmail.com, Juancarloscue@gmail.com, cbarto@gmail.com,
fiorito.nicolas@gmail.com, spalletti.mauricio@gmail.com

Resumen

Un Sistema Experto que permite la carga de la Normalización de modelos de datos para base de datos relacionales, que luego será validada hasta tercera forma normal en su diseño. Para ser utilizado en el ámbito académico dentro de las carreras de Informática, como herramienta de aprendizaje (sin supervisión docente) de los estudiantes, para la temática de Normalización de Modelos de Datos para bases de datos relacionales.

Desarrollado con la integración de tres herramientas (software) de programación: a) Vb.Net para desarrollo de la interfaz visual del software, MySql como motor de base de datos, y el lenguaje Clips de programación para el diseño de Sistemas Expertos.

Desarrollamos conocimientos sobre los temas: Ingeniería del Conocimiento, Sistemas Expertos, Ontología. Incrementamos

conocimiento sobre Normalización de Base de Datos.

Palabras clave:

Sistema Experto, Ingeniería del Conocimiento, Normalización de Base de Datos, Ontología

Contexto

El grupo de investigadores de este proyecto, está compuesto por docente de la Casa de Altos Estudios de la carrera Ingeniería en Sistemas de Información, con la colaboración e integración de un docente con nivel de experticia en la temática Sistemas Expertos (SE) de la Universidad Nacional de Córdoba. Algunos docentes de este grupo, poseen experiencia previa en el desarrollo de proyectos de investigación, encontrándose categorizados dentro del sistema para este fin de UTN y Nacional. Otros noveles, están transitando el tercer

año de trabajo en este PID, y postulando por primera vez en los sistemas de categorización, para iniciar sus carreras de investigadores.

Los docentes ya categorizados han participado en otras investigaciones, homologadas por UTN, también referidas a la temática de bases de datos, pero con otras orientaciones, estas son: PROMETEO - Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales (2010), Análisis y aplicación de metodologías para la generación de consultas complejas utilizando esquemas OLAP (2010), y administrador de Base de Datos Relacional TecnoDB (2007).

Las especialidades dentro de la informática de los docentes integrantes del PID, son las siguientes: Programación, Inteligencia Artificial, Base de Datos, Computación y Sistemas de Información.

También integran el PID, un conjunto de estudiantes avanzados de la carrera Ingeniería en Sistemas de Información.

Introducción

El inicio de este PID es motorizado por un grupo de docente de la Cátedra Gestión de Datos, de la carrera de Ingeniería en Sistemas de Información, dentro de UTN-FRC, que proponen la creación de una herramienta de utilización académica, para que los estudiantes de la carrera, puedan realizar prácticas de Normalización de estructuras de datos para Base de Datos Relacionales. Posteriormente se suman docentes de otras cátedras que encuentra la propuesta interesante. Se formula la propuesta como un PID formal dentro del ámbito UTN, que proponía: en el área de investigación desarrollar conocimiento en las temáticas de Ingeniería del Conocimiento y diseño de SE, e incrementar el conocimiento

sobre temáticas de Bases de Datos Relacionales, y como corolario de la investigación, el desarrollo de un software con las siguientes características: a) un generador de estructuras de datos Normalizas, para base de datos relacionales, a partir de la declaración de dominios, b) un validador de la normalización de estructuras de datos para base de datos relacionales, hasta tercera forma normal (3FN) [1,2,3,4], suministrado por un usuario, con la característica de emitir opinión respecto de las partes de la estructura de datos, donde el SE encontrara errores de diseño.

Se debió realizar una tarea intensa de búsqueda de información, bibliografía y webgrafía referida al diseño del SE [6,9,15], también sobre Ingeniería del Conocimiento(IC) [5,7], para producir un aprendizaje que nos permitiera iniciarnos dentro de este campo de la informática donde ninguno de los integrantes del grupo, había desarrollado conocimiento y experiencias. Hallamos amplias dificultades en este proceso de recopilación de bibliografía, y como consecuencia, se decidió invitar al Ing. Carlos Bartó de Universidad Nacional de Córdoba a participar en el PID, quien se destaca por su nivel de experticia en las temáticas antes mencionadas. Esta incorporación trajo luz a nuestro proyecto, saliendo del estancamiento que naturalmente se estaba produciendo por la falta de acceso a material necesarios. Culminada nuestra capacitación, nos encontramos con el primer año de proyecto avanzado, con una visión esclarecidas y renovada de nuestros objetivos, con una mirada realizable del producto que debíamos desarrollar. Como consecuencia de esto la definición del PID cambió radicalmente. Determinamos que el dominio del problema era demasiado amplio para ser tratado por un único SE, para que éste fuera realizable,

se debía subdividir la problemática a dominios más acotados o restringidos, esto implicaba el rediseño conceptual de nuestros objetivos y por ende del desarrollo del software. Se concluyó que se debía abordar una solución por vez, por ello nos enfocamos en el desarrollo del Sistema Validador de estructuras de datos, y dejamos para un siguiente proyecto el Generador de Modelos de Datos en función de dominios suministrados, que fue la idea que dio nombre al PID original.

En el inicio del segundo año se planificó una nueva estrategia para darle continuidad al proyecto. Se decide el desarrollo de una interfaz visual que le permita a un usuario, la carga de una estructura de datos (modelo de datos), normalizada a tercera forma normal para base de datos relacionales, que representa la solución de un problema de almacenamiento.

Nos encontramos en el tercer año de desarrollo del PID, ya hemos culminado la interfaz visual de la herramienta, la que hemos desarrollado Vb.Net 2010 [13,16]. Esta parte del producto permitirá al usuario plasmar un modelo de datos [1], como dijimos anteriormente, de la misma forma en que se realiza un ejercicio de aprendizaje de normalización [2,4] (para base de datos relacionales) en papel. Esta interfaz le brindará los recursos suficientes para declarar cada una de las entidades componente de un modelo o estructura de datos, mencionar la solución con nombre propio, y poder almacenar la misma en una base de datos MySql [22]. Permitiéndole también trabajar en un entorno de múltiples modelos (resultados) simultáneos.

El desafío del tercer año y último, será el embeber en la interfaz visual, un módulo que permita ejecutar la programación del SE, el que se escribirá en el lenguaje Clips [6,10,21]. Ya

disponemos de ese módulo (DLL) que está compilado en C#[17], preparado para montarse en un proyecto .Net, proveniente del sitio CLIPS.NET [23], que es sugerido en el sitio oficial de Clips [25]. Hemos realizado algunas pruebas y las mismas han sido satisfactorias.

Sin dudas aquí, en este punto del desarrollo del PID, nos encontramos con el dilema de cómo definir el conocimiento. Dentro de los sistemas de producción, la definición típica sostiene que la base de conocimiento está representada por reglas [6]. Nosotros entendemos que las reglas son el conocimiento de cómo hacer (o procedimental), pero que existe otro conocimiento que podríamos denominar ontológico, que es la acumulación de concepto definitorios de una misma tipificación. Por ejemplo: El cómo hacer, que representa en el experto humano, las tareas sistematizadas para lograr que un modelo de base de datos se encuentre en primera forma normal, en el SE (programación Clips) implicarán la base del conocimiento necesaria, representada por reglas, para lograr el mismo objetivo. Pero el experto humano, dispone de un conocimiento acumulado por la experiencia, de cómo mencionar entidades y atributos, que se incrementa con el tiempo, y que no pertenece al campo de las Formas Normales, que constituyen ontologías, pues para un mismo concepto dispone de una variedad aceptable de expresiones que lo representan. Estas ontologías forman parte del conocimiento, pero no es conveniente que se las incluya en la base del conocimiento (reglas), pues es un conjunto de datos volátil. La mejor alternativa es que las ontologías se encuentren almacenadas en una base de datos, y puedan ser corregidas y validadas por un ontólogo, y requeridas por el SE en cada análisis. Estas ontologías deberán

encontrar representación en los tres sistemas actuantes: la base de datos, la aplicación visual y el SE embebido.

El producto es un SE en donde interactúan tres componentes (sistemas) asociados para lograr el efecto esperado. Primero una *interfaz visual* que permite declarar un modelo de datos (normalización) en forma visual, similar a como lo escribiría el usuario en papel, con una representación en pantalla aproximada. Esto requiere de un visualizador especializado de entidades, que le permite al usuario ver las entidades declaradas, con sus atributos, clave primaria y foráneas. El objeto que permite *visualizar una entidad* es flotante dentro del escritorio de normalización, pudiendo el usuario agregar la cantidad que fueran necesarias en el diseño de la estructura de datos. El escritorio de normalización es de pantalla completa, y se encuentra manejado por un PageFrame (manejador de páginas) que permite N Page (página) dentro de sí, este mecanismo faculta al usuario a declarar un modelo de datos por cada Page. El *editor de entidades*, es la herramienta que le permite al usuario dar contenido a una entidad, es un formulario especializado que se activa cada vez que se desea editar una entidad desde el objeto visualizador que la contiene, tiene todos los recursos para definir entidad completamente, su nombre, su colección de atributos, la clave primaria, y las claves foráneas, una vez terminado el proceso de edición, se formatea el conjunto de datos, y se los transfiere al *visualizador de entidades* desde donde se activó el editor. Cada estructura de datos, puede ser almacenada dentro de una base de datos en el motor MySQL, con las siguientes características: a) se guarda la definición (nombre propios) de la estructura (modelo), asociado a un usuario registrado, y a una fecha, se procede a guardar las entidades definidas

pertenecientes a ese modelo, agregando a la información de la entidad, su ubicación dentro del editor (coordenadas de localización), luego se almacena la definición de cada atributo, por último la clave primaria, y las foráneas. Ya que se grabó para cada modelo, la localización de visualización en el escritorio de normalización, de cada entidad, esto permitirá la posibilidad de ubicarlas en el mismo lugar dentro del escritorio de normalización a cada una de ellas, cuando vuelva a ser cargado los modelos de datos desde la base de datos, para ser reutilizado. Esta forma de mantener el diseño esquemático visual que el usuario ha entendido como correcto, está pensado para dar continuidad, con el propósito de no generar una relectura del diseño realizado hasta el momento, para continuarlos. Siempre tratando de que el escritorio de normalización de los una mejora sustancial respecto de los métodos de papel y lápiz.

También existe la funcionalidad de borrar todo el modelo físicamente en la base de datos, o borrar todo el contenido del page actual.

Una vez declarado todo el modelo, el usuario puede enviar a validar su diseño, esto activa la parte del SE que contiene programación Clips. El sistema visual convierte toda la información del modelo de datos, en un conjunto de patrones que representa la misma información bajo otro paradigma. Se transmite este conjunto de patrones al SE, en otra acción la interfaz visual lee el paquete de conocimiento que denominamos ontológico desde la base de datos, es transformado en patrones compatibles con la programación Clips, y transferidos al SE, por último se dispara la ejecución del SE (motor de inferencia).

Hemos comprendido que la definición de una serie de ontologías, que ajusten: la representación del conocimiento, del

modelo a ser analizado, y de los resultados obtenidos, es de fundamental importancia. Estas ontologías garantizan que la información pueda ser vista desde cualquiera de los sistemas interactuantes con el mismo significado.

Líneas de Investigación, Desarrollo e Innovación

Inicialmente el PID propone investigar sobre el diseño de Sistema Expertos programados en Clips, Ingeniería del Conocimiento, y logra un incremento en conceptos de Normalización de Base de Datos Relacionales. En el desarrollo del mismo, se incorporaron líneas de conocimiento no previstos sobre CommonKADS [7,8], y Ontología [18,22].

Se están logrando los objetivos de desarrollo, avanzando fuertemente en la construcción de un Sistema Experto, montado sobre un interfaz visual, con las siguientes características: a) la carga de un modelo solución (estructura de datos) a un problema, que debiera estar normalizado en 3FN, b) la validación de esta estructura de datos, c) emisión de opinión: c.1) en caso de no encontrar anomalías expresar que la estructura es satisfactoria, c.2) por lo contrario, emitir sugerencias que le permitan al usuario realizar las correcciones pertinentes para que su modelo alcance un estado de 3FN, c.3) también la interacción entre SE y el usuario para aclarar situaciones dudosas.

Respecto de los conocimientos no previstos: CommonKADS es una metodología para documentar SE, que está diseñada específicamente para este propósito. Los conceptos de Ontología fueron imprescindibles para poder trabajar en forma sistémica con el conocimiento de cómo mencionar los componentes de un modelo de datos,

basado en la experiencia del experto humano.

Resultados y Objetivos

Respecto de los resultados, hasta este momento tenemos desarrollada la Interfaz visual que permite capturar un modelo de datos. También la parte de la base de datos que permite el almacenamiento de los modelos capturados. En el año en curso, nos abocaremos al desarrollo del SE junto a los módulos de la interfaz visual que compatibilizan, el conocimiento ontológico y el del modelo a ser validado, junto a la estandarización de mensajes y sugerencias que el sistema emitirá.

Respecto de los objetivos, los originales del proyecto ya fueron expuestos en wicc 2013, cuando fue presentado como un potencial en el inicio del PID. Ahora nos gustaría dejar plasmado en este apartado la siguiente reflexión sobre objetivos: entendiendo a estos como el futuro por venir, sin dudas nos queda la total convicción de difundir la experiencia lograda, fundamentalmente la relacionada al cambio de pensamiento, que debieran propiciar los docentes universitarios de carreras asociadas a la tecnología (en principio), en relación a qué, productos tecnológicos de cualquier tipo, desarrollados en el seno de la universidad, pueden cumplir el rol de apoyo a la enseñanza, para mejorar las posibilidades de entrenamiento y prácticas de estudiantes, brindando recursos adicionales extra áulicos para su formación.

Formación de Recursos Humanos

El equipo de trabajo está constituido por: a) un director de proyecto, b) un codirector, c) cuatro docentes investigadores, d) seis alumnos becarios,

con la siguiente distribución en el transcurso de los años: d.1) primer año 4 becarios, d.2) segundo año 2 becarios, d.3) tercer año (el actual) 2 becarios, los becarios del segundo y tercer año son los mismos.

En el segundo año del PID un estudiante de la carrera Ingeniería en Sistemas de Información realizó sus *prácticas supervisadas* de 200 horas de trabajo en el seno del proyecto. Las prácticas supervisadas tiene como objetivo, intensificar la formación práctica de los alumnos, desarrollando una formación científica - técnica actualizada y adecuada a las necesidades de un medio en continua evolución [23].

Referencias

- [1] Elmasri, R. y Navathe, S. - Fundamentos de Sistemas de Bases de Datos-5ta Edición – Edit. Pearson- Estados Unidos - 2007. ISBN: 978-84-7829-085-7.
- [2] Date, Christopher - Introducción a los Sistemas de Bases de Datos - Volumen 1 - Quinta Edición - Edit. Addison Wesley Iberoamericana – Estados Unidos - 1993 - ISBN: 0-201-51859-7.
- [3] Silberschatz y otros- Fundamentos de Bases de Datos - Quinta Edición- Edit. Mc Graw Hill- Estados Unidos - 2006 - ISBN: 84-481-4644-1.
- [4] Reinoso, E.; Maldonado, C.; Muñoz, R.; Damiano, L.; Abrutsky, M. - Bases de Datos - Edit. AlfaOmega Editores - Argentina - 2012- ISBN: 978-987-1609-31-4.
- [5] Alonso Betanzos, Amparo et. al. - Ingeniería del Conocimiento. Aspectos Metodológicos - Editorial Pearson S.A. - Madrid - 2004 - ISBN 84-205-4192-3.
- [6] Giarratano, Joseph - Riley, Gary - Sistemas Expertos Principios y Programación - Editorial Cengage Learning / Thomson Internacional, 2005 - ISBN 9789706860590
- [7] Pajares Martinsanz, Gonzalo y Santos Peñas, Matilde - Inteligencia Artificial e Ingeniería del Conocimiento - Editorial RA-MA - Madrid - 2007 - ISBN 84-7897-676-0.
- [8] Schreiber, August et. al. - Knowledge Engineering and Management. The CommonKADS Methodology - Editorial Massachusetts Institute of Technology - Estados Unidos - 2000 - ISBN 0-262-19300-0.
- [9] Alanis Barrera, Ricardo- Sistemas Expertos e Inteligencia Artificial CLIPS - Burgos CyL., Febrero de 2005 - Universidad de Burgos: <http://es.scribd.com/doc/6732509/Clips>. Accedido en junio de 2013.
- [10] Calvo Cuenca, Antonio - Programación en lenguaje Clips - Editorial Universitaria Ramón Areces, 2008 - ISBN 9788480048798.
- [11] García Martínez, Ramon - Sistemas Autónomos - Aprendizaje Automático - Editorial Nueva Librería - Argentina - 1997 - ISBN 950-9088-84-6
- [12] Blanco Roberto; Casado, María Luisa; Cervera, David; Días de Prado, Fausto; Gómez Arias, José Javier; Martín, Francisco José; Martínez Martínez, José Miguel; Ramos, María José - Tecnología Investigación, innovación y buenas prácticas - Aprendizaje Automático - Editorial Graó - España - 2010 - ISBN 978-84-7827-995-1
- [13] Wroblewski, Luke - Usabilidad y diseño de Formularios - 2011 - <http://www.slideshare.net/interactionpatterns.org/best-practices-for-form-design-presentation> Disponible: Agosto 2014
- [14] García Serrano, Alberto - Inteligencia Artificial - Fundamentos, práctica y aplicaciones - Editorial Alfaomega - México - 2012 - ISBN 9788493945022

- [15] García Martínez - Britos - Ingeniería de Sistemas Expertos - Editorial Nueva Librería - Bs.As. Argentina - 2004 - ISBN 9789871104154
- [16] Ceballos, Francisco Javier - Visual Basic .NET - Curso de Programación - Editorial Alfaomega - Rama - México - 2008 - ISBN: 9701513703
- [17] Ceballos, Francisco Javier - Enciclopedia de Microsoft - Visual C# - Editorial Alfaomega - México - 2007 - ISBN: 9789701514252
- [18] Mario Bunge – Tratado de Filosofía – Ontología I – Editorial Gedisa – Barcelona 2011 – ISBN: 978-84-9784-196-2
- [19] Mario Bunge - Tratado de Filosofía – Ontología II – Editorial Gedisa – ISBN: 978-84-9784-197-9
- [20] Paul DuBois - Edición Especial MySQL - Editorial Prentice Hall - ISBN: 84-205-3299-1
- [21] Gary Riley - Giarratano Joseph - Otros - Clips Guía de Programación Básica , Clips Guía de Programación Avanzada, Otros Títulos - <http://clipsrules.sourceforge.net/> (marzo 2015)
- [22] María Jesús Lamarca Lapuente - Ontologías - <http://www.hipertexto.info/documentos/ontologias.htm> (marzo 2015)
- [23] Sitio Clips.Net - <https://clips.codeplex.com> (marzo 2015)
- [24] UTN -FRC - Prácticas Supervisadas <http://www.institucional.frc.utn.edu.ar/sistemas/Areas/Alumnos/PracticasSupervisadas.asp> (marzo 2015)
- [25] Sitio Oficial de Clips - <http://clipsrules.sourceforge.net/>